EATCS Monographs on Theoretical Computer Science

O. Watanabe (Ed.)

# Kolmogorov Complexity and Computational Complexity

Springer-Verlag

EATCS
Monographs on Theoretical Computer Science



Osamu Watanabe (Ed.)

# Kolmogorov Complexity and Computational Complexity

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

*Volume Editor*

Prof. Dr. Osamu Watanabe
Department of Computer Science
Tokyo Institute of Technology
Meguro-ku, Ookayama
Tokyo 152, Japan

*Editors*

Prof. Dr. Wilfried Brauer
Institut für Informatik, Technische Universität München
Arcisstrasse 21, W-8000 München 2

Prof. Dr. Grzegorz Rozenberg
Institute of Applied Mathematics and Computer Science
University of Leiden, Niels-Bohr-Weg 1, P. O. Box 9512
2300 RA Leiden, The Netherlands

Prof. Dr. Arto Salomaa
The Academy of Finland
Department of Mathematics, University of Turku
SF-20 500 Turku, Finland

ISBN 3-540-55840-3 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-55840-3 Springer-Verlag New York Berlin Heidelberg

Library of Congress Cataloging-in-Publication Data
Watanabe, Osamu, 1958- Kolmogorov complexity and Computational Complexity / Osamu Watanabe. p. c. - (EATCS monographs on theoretical computer science) "In March 1990, the Symposium on Theory and Application of Minimal Length Encoding was held at Stanford University as part of the AAAI 1990 spring symposium series" - Galley. Includes bibliographical references and index.
ISBN 0-387-55840-3 (N.Y.)
1. Kolmogorov complexity - Congresses. 2. Computational complexity - Congresses. I. Title. II. Series. QA267.7.W38 1992 511.3 - dc20 92-26373



Typesetting: Camera ready by editor using the TEX macro package from Springer-Verlag.
45/3140 - 5 4 3 2 1 0 - Printed on acid-free paper

# Preface

The mathematical theory of computation has given rise to two important approaches to the informal notion of "complexity": *Kolmogorov complexity*, usually a complexity measure for a single object such as a string, a sequence etc., measures the amount of information necessary to describe the object. *Computational complexity*, usually a complexity measure for a set of objects, measures the compuational resources necessary to recognize or produce elements of the set. The relation between these two complexity measures has been considered for more than two decades, and may interesting and deep observations have been obtained.

In March 1990, the Symposium on Theory and Application of Minimal-Length Encoding was held at Stanford University as a part of the AAAI 1990 Spring Symposium Series. Some sessions of the symposium were dedicated to Kolmogorov complexity and its relations to the computational complexity theory, and excellent expository talks were given there. Feeling that, due to the importance of the material, some way should be found to share these talks with researchers in the computer science community, I asked the speakers of those sessions to write survey papers based on their talks in the symposium. In response, five speakers from the sessions contributed the papers which appear in this book.

In this book, the main topic is Kolmogorov complexity and its relations to the structure of complexity classes. As I explain in the Introduction, each paper discusses a different type of Kolmogorov complexity, and each paper uses a different viewpoint in developing a relationship between Kolmogorov complexity and computational complexity. Thus, this book provides a good overview of current research on Kolmogorov complexity in structural complexity theory.

I whish to thank Dr. Edwin Pednault, the Chair of the Symposium, for having organized the interesting sessions from which this book originated. Each paper was reviewed by some outside reviewer as well as by fellow authors. I would like to thank the outside reviewers, Professor José Balcázar, Professor Kojiro Kobayashi, and Professor Keri Ko, for their constructive comments.

*Osamu Watanabe*
*May 1992*

# Complexity and Entropy: An Introduction to the Theory of Kolmogorov Complexity *

*Vladimir A. Uspensky*

Department of Mathematical Logic
Faculty of Mechanics and Mathematics
Moscow Lomonosov University
V-234 Moscow, GSP-3, 119899 Russia
uspensky@globlab.msk.su

**Contents.**



## 1 Complexity, Entropy, and Randomness

Things can be large or small, and their size (the length or the volume or the weight or so on) can be measured by a number. Besides, things can be simple or complex, and their complexity can also be measured by a number. I do not know to whom we are indebted for measuring sizes by numbers. It was Andrei Kolmogorov [Kol65] who proposed to measure the complexity of a thing by a natural number (i.e., a non-negative integer), and he developed the rudiments of the theory.

Complexity of things (as opposed to the complexity of processes, e.g., of computational processes) took the name *descriptional complexity*, or *Kolmogorov complexity*. As will be seen here, in appropriate cases one may say "entropy" instead of "complexity".

* Preparation of this paper was supported in part by the Institute of New Technologies at Moscow.

Thus we assume that there is a set $Y$ of things, or objects, $y$'s, and a total function "complexity of $y$" defined on $Y$. That function will be denoted by Compl and its possible values are $0, 1, 2, 3, ..., n, ..., \infty$. So the function Compl is a total function from $Y$ to $\mathbb{N} \cup \{\infty\}$. We do not put any further restrictions on Compl, but take it on an intuitive level as a measure of complexity, or a complexity function, or, shorter, a complexity.

Let $\text{Compl}_1$ and $\text{Compl}_2$ be two measures of complexity. Let us say that $\text{Compl}_1$ is *not worse* than $\text{Compl}_2$ if

$$\text{Compl}_1(y) \underset{+}{\leq} \text{Compl}_2(y).$$

**Explanation.** The notation $A(y) \underset{+}{\leq} B(y)$ means that for some constant $c$ not depending on $y$ and for all $y$, $A(y) \leq B(y) + c$ holds.

Let $\mathcal{Z}$ be some class of complexities, or (that is the same) of measures of complexity. Let $\text{Compl}_0$, belonging to $\mathcal{Z}$, be not worse than any complexity belonging to $\mathcal{Z}$. Then $\text{Compl}_0$ is called *optimal* in the class $\mathcal{Z}$. So a way of measuring complexity is called optimal if it gives, roughly speaking, the shortest complexities of things. Of course, a class of complexities may have no optimal one.

Any optimal complexity is called an *entropy*. It is possible that a class $\mathcal{Z}$ has several entropies, but any two entropies, $\text{Ent}_1$ and $\text{Ent}_2$, fulfill the following condition:

$$\text{Ent}_1(y) \underset{+}{=} \text{Ent}_2(y)$$

**Explanation.** The notation $A(y) \underset{+}{=} B(y)$ means that $|A(y) - B(y)| \underset{+}{\leq} 0$, or $A(y) \underset{+}{\leq} B(y)$ *and* $B(y) \underset{+}{\leq} A(y)$.

**Important Remark.** There is no semantic problem when one speaks about *an* entropy related to some class of complexities. But in the theory of Kolmogorov complexity it is usual to speak about *the* entropy and even to denote it by a special notation. What does it mean? Here we have an *abus de langage* (after N.Bourbaki). Speaking about *the* entropy related to some class, one speaks in fact about *an arbitrary* entropy of that class. And the notation denotes any of such entropies. Of course, our statements must be invariant and do not change their truth value when a particular entropy changes to another one but still belonging to the same class. But we must be cautious. Let $\mathcal{V}$ and $\mathcal{W}$ be two classes of complexity functions, and let $K$ be *the* entropy related to $\mathcal{V}$ and $L$ be *the* entropy related to $\mathcal{W}$. In fact, $K$ and $L$ denote two families of entropies, or, it is better to say, any entropies of those two families. When we write $K(y) \underset{+}{\leq} L(y)$, we suppose that this relation $\underset{+}{\leq}$ holds for any particular entropy denoted by $K$ and any particular entropy denoted by $L$ (so there is an additive constant hidden in this relation depending on the choices of particular representatives of $K$ and

$L$). But when we declare that $K$ and $L$ coincide (are *the same* entropy), we do not want to express the opinion that any entropy denoted by $K$ coincides with any entropy denoted by $L$. That is, we understand the coincidence statement in the following way: for any of the entropies $K$ and any of the entropies $L$, there exists a constant $c$ such that $|K(y) - L(y)| < c$ for all $y$.

**Terminological Remark.** In literature on Kolmogorov complexity, the term "complexity" (synonymous with "complexity function" and "measure of complexity") is most often used in the sense of the term "entropy". But we make the distinction between those two terms: entropy is an optimal complexity.

As has already been said, there may be no entropy among complexity functions belonging to a class $\mathcal{Z}$. An important property of a class $\mathcal{Z}$ is that of having an entropy. In such case, we say that the *Solomonoff–Kolmogorov Theorem* holds for $\mathcal{Z}$.

There exist several important classes of complexities that contain entropies. And among those entropies, there are ones of special interest — namely, those entropies that can be used for a definition of randomness. Kolmogorov has proposed the following definition of randomness for an infinite binary sequence $a_1a_2a_3\cdots a_n\cdots$: the sequence is called *random*, or, more exactly, *Kolmogorov random*, if

$$\mathrm{Ent}(a_1\cdots a_n) \underset{\oplus}{\geq} n,$$

where Ent is an entropy. Of course, the choice of Ent is to be specified. Not every sort of entropy goes to a "good" definition of randomness; a definition by Kolmogorov scheme is regarded as "good" if the class of Kolmogorov random sequences sprung up by that definition coincides with the class of sequences that are random in the sense of Martin-Löf (or are *typical sequences*; see [KU87a] and [KU87b] and [Mar66]).

To sum up: in order to define an entropy, one must define an appropriate class of complexities and show that the Solomonoff–Kolmogorov Theorem holds for that class.

## 1.1 Generation of Complexities by Means of an Encoding Procedure

The idea (due to Kolmogorov) is very simple. There are objects and there are descriptions (encodings) of objects, and the complexity of an object is the minimal size of its description.

In more detail, there is a set $Y$ of objects $y$, and a set $X$ of descriptions (names, encodings) $x$. There is a volume function $\ell$ defined on $X$; that $\ell$ is a function from $X$ to $\mathbb{N}$. A *mode of description*, or a *description mode*, is an arbitrary set $E \subseteq X \times Y$. If $\langle x, y\rangle \in E$, then $x$ is called a *description* (a *name*, an *encoding*) *of* $y$ *with respect to* $E$. Thus an object $y$ may have many descriptions and a description may serve as a description for many objects.

The *complexity of* $y$ *with respect to* a description mode $E$ is defined as follows:

$$\mathrm{Compl}_E(y) = \min\{\ell(x) : \langle x, y\rangle \in E\}.$$

We make the convention that $\mathrm{Compl}_E(y) = \infty$ if there is no $x$ such that $\langle x, y\rangle \in E$.

Let $\mathcal{E}$ be a class of modes of description. Each mode $E \in \mathcal{E}$ gives the corresponding complexity function $\mathrm{Compl}_E$. Then there arises the class $\mathcal{Z} = \mathcal{Z}(\mathcal{E})$ of all complexity functions related to the modes of $\mathcal{E}$, and one may ask whether the class $\mathcal{Z}$ contains an optimal function, or an entropy. If such an optimal function exists, then it corresponds to some description mode which is also called *optimal.*

Until now we have not imposed any restrictions on $X, Y, E$. It is reasonable to assume that $X$ and $Y$ consist of *constructive objects*, and $E$ is a *generable* set (in the sense of Post) and, consequently, is a recursively *enumerable* set.

In the following exposition we shall restrict ourselves with the following simple case: Both $X$ and $Y$ are $\Xi$, where $\Xi$ is the set of all binary words, or finite binary sequences. The volume function $\ell$ is defined to be $\ell(\xi) = |\xi|$ for every $\xi \in \Xi$, where $|\xi|$ is the length of $\xi$.

### 1.2 Two Symmetric Relations and Four Entropies

Our task is to define—in a reasonable way—a class $\mathcal{E}$ of modes of description as a class of subsets of $\Xi \times \Xi$. Having this goal in mind, we define a binary relation on $\Xi$ which we shall call *the concordance relation*: $u_1$ and $u_2$ are called *concordant* if they have a mutual continuation, i.e., if there are $t_1, t_2 \in \Xi$ such that $u_1 t_1 = u_2 t_2$. This concordance relation will be denoted by $\gamma$.

Thus we have two natural binary relations on $\Xi$: the equality $=$ and the concordance $\gamma$. Both are symmetric and decidable (see Explanation in Sect.1.3).

Let $\alpha$ and $\beta$ be two binary relation on $\Xi$. We say that a mode of description $E \subseteq \Xi \times \Xi$ *fulfills the* $(\alpha, \beta)$*-property* if for every $x_1, x_2, y_1, y_2 \in \Xi$,

$$\langle x_1, y_1\rangle \in E \wedge \langle x_2, y_2\rangle \in E \wedge x_1 \alpha x_2 \;\Rightarrow\; y_1 \beta y_2.$$

Let us consider the class $\mathcal{E} = \mathcal{E}(\alpha, \beta)$ of all recursively enumerable modes of description that fulfill $(\alpha, \beta)$-property and the related class $\mathcal{Z}^{\alpha,\beta} = \mathcal{Z}(\mathcal{E})$ of complexity functions. If the class $\mathcal{Z}^{\alpha,\beta}$ contains an optimal complexity function, that complexity function will be called $(\alpha, \beta)$*-entropy.*

Now move from variable $\alpha$ and $\beta$ to constants $=$ and $\gamma$. Taking $=$ or $\gamma$ as $\alpha$ or $\beta$, we obtain four classes of complexity functions: $\mathcal{Z}^{=,=}$, $\mathcal{Z}^{=,\gamma}$, $\mathcal{Z}^{\gamma,=}$, $\mathcal{Z}^{\gamma,\gamma}$. For each of these four classes, the Solomonoff–Kolmogorov theorem is valid, so we have four entropies:

1. $(=, =)$-entropy, or $\mathbb{N}\mathbb{N}$-entropy,
2. $(=, \gamma)$-entropy, or $\mathbb{N}\Xi$-entropy,
3. $(\gamma, =)$-entropy, or $\Xi\mathbb{N}$-entropy, and
4. $(\gamma, \gamma)$-entropy, or $\Xi\Xi$-entropy.

*Note.* The notations $\mathbb{NN}$, etc., have the following origin. In [US81] and [US87], the notation "$\Xi$" had the following meaning: the set of all binary words being considered together with relation $\gamma$. In place of the set of all binary words with the relation $=$ on that set, the set $\mathbb{N}$ of all natural numbers with the relation $=$ and the volume function $l(x) = \lfloor \log_2(x+1) \rfloor$ was considered. This volume function is induced by the following 1-1 correspondence between $\mathbb{N}$ and $\Xi$: zero $\sim \Lambda$, one $\sim 0$, two $\sim 1$, three $\sim 00$, four $\sim 01$, five $\sim 10$, and so on.

### 1.3 Two Approximation Spaces and Four Entropies

There is another way to come to the four basic entropies of Sect.1.2.

Any set of constructive objects with a decidable partial ordering defined on that set will be called an *approximation space.*

**Explanation.** The term "decidable" means that there is an algorithm to decide for any $x'$ and $x''$, whether $x' \leq x''$ or not.

On an intuitive level, the elements of an approximation space can be taken as informations, and $x' \leq x''$ means that the information $x''$ is a refinement of the information $x'$ (and hence $x''$ is closer than $x'$ to some limit value to which both $x'$ and $x''$ serve as approximations).

To develop a more attractive theory of approximation spaces, especially with the intention to apply this theory to an advanced theory of Kolmogorov complexity, one needs to include some additional requirements into the definition of an approximation space. For our goals, however, it suffices to have a decidable partial ordering. Moreover, only two approximation spaces will be considered: the bunch $\mathbb{IB}$ and the tree $\mathbf{T}$. Their definitions follow immediately.

- The *bunch* $\mathbb{IB}$: The set of objects is $\Xi$, and the partial ordering $\leq$ is $=$, i.e., $u \leq w$ iff $u = w$.
- The *tree* $\mathbf{T}$: The set of objects is $\Xi$. The partial ordering $\leq$ is defined as follows: $u \leq w$ iff $u$ is a prefix of $w$ (and $w$ is a continuation of $u$), i.e., $\exists v[\, uv = w\,]$.

Let $X$ and $Y$ be two approximation spaces. The spaces $X$ and $Y$ will be treated, respectively, as the space of descriptions (names, encodings) and as the space of the objects described (named, encoded). Our near goal is to define the class $\mathcal{E}$ of acceptable description modes $E \subseteq X \times Y$.

We impose on $E$ the following three requirements:

1. if $\langle x, y\rangle \in E$ and $x' \geq x$, then $\langle x', y\rangle \in E$,
2. if $\langle x, y\rangle \in E$ and $y' \leq y$, then $\langle x, y'\rangle \in E$, and
3. if $\langle x, y_1\rangle \in E$ and $\langle x, y_2\rangle \in E$, then there exists a $y$ that $\langle x, y\rangle \in E$, $y_1 \leq y$, and $y_2 \leq y$.

Hence, the only cause of the existence of two different objects having the same description is the execution of the first requirement.

A description mode is called *acceptable* if it is (recursively) enumerable and fulfills all the above requirements.

If one wishes to relate a complexity function with any description mode, one needs to introduce a volume function $\ell$ defined on $X$. Here we are interested in the cases $X = \mathbb{B}$ and $X = \mathbf{T}$ only. In both these cases, we put $\ell(x) = |x|$ where $|x|$ is the length of $x$.

Now let us fix approximation spaces $X$ and $Y$, and let us consider the class of all acceptable description modes and the corresponding class of complexities. Let us ask whether the Solomonoff–Kolmogorov theorem holds for that class, and, if it does hold, then the related entropy will be called $XY$ *entropy*.

It turns out that the Solomonoff–Kolmogorov theorem is valid for four cases when $X$ and $Y$ is respectively $\mathbb{B}$ or $\mathbf{T}$:

1. For $X = \mathbb{B}$ and $Y = \mathbb{B}$, we have $\mathbb{B}\mathbb{B}$-entropy,
2. For $X = \mathbb{B}$ and $Y = \mathbf{T}$, we have $\mathbb{B}\mathbf{T}$-entropy,
3. For $X = \mathbf{T}$ and $Y = \mathbb{B}$, we have $\mathbf{T}\mathbb{B}$-entropy, and
4. For $X = \mathbf{T}$ and $Y = \mathbf{T}$, we have $\mathbf{T}\mathbf{T}$-entropy.

It is easy to see that $\mathbb{B}\mathbb{B}$-, $\mathbb{B}\mathbf{T}$-, $\mathbf{T}\mathbb{B}$-, $\mathbf{T}\mathbf{T}$-entropy respectively coincides with $(=,=)$-, $(=,\gamma)$-, $(\gamma,=)$-, $(\gamma,\gamma)$-entropy of Sect.1.2. Speaking on the coincidence, take into account the Important Remark of Sect.1.

### 1.4 The Ordering of the Four Entropies

Now we have four entropies, and any two of them, $A$ and $B$, do not coincide; which means that the assertion $A(y) \underset{\Pi}{=} B(y)$ is not valid. Let us write $A < B$ if $A(y) \underset{\Pi}{\leq} B(y)$ but not vice versa. Then there is a partial ordering on the set of four entropies. That ordering can be shown by the following picture; Fig.1.

The picture is directed from bottom to top. That is, it shows that

$$\mathbb{B}\mathbf{T} < \mathbb{B}\mathbb{B},\ \mathbb{B}\mathbf{T} < \mathbf{T}\mathbf{T},\ \mathbb{B}\mathbb{B} < \mathbf{T}\mathbb{B},\ \mathbf{T}\mathbf{T} < \mathbf{T}\mathbb{B},$$

and, of course,

$$\mathbb{B}\mathbf{T} < \mathbf{T}\mathbb{B}.$$

On the other hand,

$$\text{neither } \mathbb{B}\mathbb{B} < \mathbf{T}\mathbf{T} \text{ nor } \mathbf{T}\mathbf{T} < \mathbb{B}\mathbb{B}.$$

### 1.5 Encoding-Free Generation of Complexities and Entropies

It turns out that the four entropies of Sect.1.4 admit an encoding-free definition with no use of such terms as "descriptions", "names", or "encodings".

As before and always, an entropy is defined as an optimal complexity function for some class $\mathcal{Z}$ of complexity functions; and all members of $\mathcal{Z}$ are functions from $Y$ to $\mathbb{N} \cup \{\infty\}$. So our goal is to describe appropriate classes $\mathcal{Z}$.

Having this goal in mind, let us introduce two conditions, $\mathbf{C}$ and $\Sigma$, which could be imposed on a function f: $Y \to \mathbb{N} \cup \{\infty\}$.

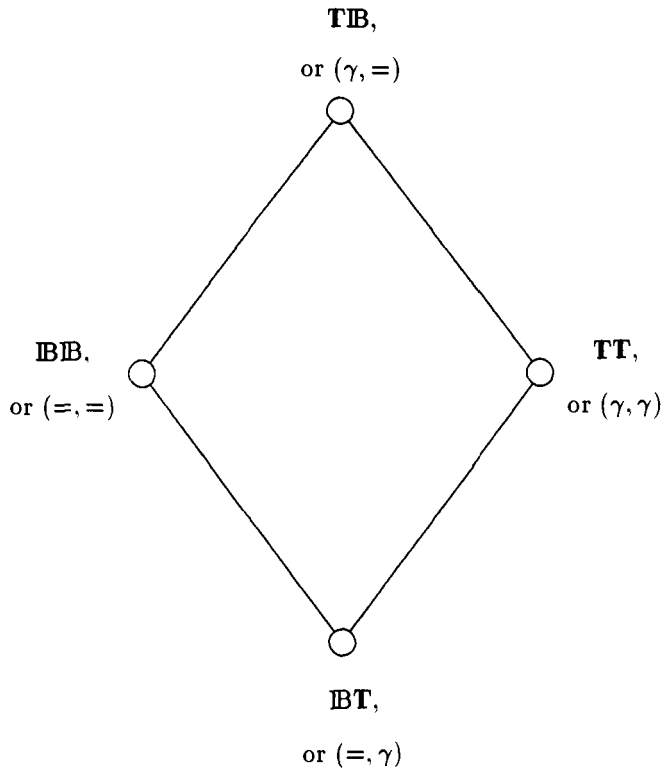


**Fig. 1.** The ordering of the four entropies: **TIB**, **IBIB**, **TT**, **IBT**

**Condition C** (of Cardinality of a set). Let $n \in \mathbf{N}$, and let $M \subseteq Y$ be an arbitrary set such that

1. any two elements of $M$ are non-comparable, and
2. $M \subseteq f^{-1}(n)$.

Then the cardinality of $M$ is less than or equal to $2^n$.

**Condition $\Sigma$** (of summation of a series). Let $M \subseteq Y$ be an arbitrary set such that any two elements of $M$ are non-comparable. Then

$$\sum_{y \in M} 2^{-f(y)} \leq 1.$$

**Explanation.** Elements $y_1$ and $y_2$ are *non-comparable* if neither $y_1 \leq y_2$ nor $y_2 \leq y_1$.

Thus, an arbitrary function $f$ may or may not satisfy Condition **C** or Condition $\Sigma$. And it is easy to see that Condition $\Sigma$ implies Condition **C**.

Further, a definition of "enumerability from above" is to appear. A function $f : Y \to \mathbb{N} \cup \{\infty\}$ is called *enumerable from above* if the set $\{ \langle y, n \rangle : y \in Y, n \in \mathbb{N}, f(y) \leq n \}$ is enumerable, that is, recursively enumerable.

Let us denote by $\mathcal{Z}(\square, Y)$ the class of all functions from $Y$ to $\mathbb{N} \cup \{\infty\}$ that are enumerable from above and satisfy the condition $\square$, where $\square$ is either $\mathbf{C}$ or $\Sigma$. Any element of $\mathcal{Z}(\square, Y)$ may be called a $\square$-*acceptable complexity*. Hence, we have four classes of acceptable complexities: $\mathcal{Z}(\mathbf{C}, \mathbb{B})$, $\mathcal{Z}(\mathbf{C}, \mathbf{T})$, $\mathcal{Z}(\Sigma, \mathbb{B})$, and $\mathcal{Z}(\Sigma, \mathbf{T})$. For each of these four classes, there holds the Solomonoff–Kolmogorov theorem.

Thus, there are four entropies: $\mathbf{C}\mathbb{B}$-entropy, $\mathbf{CT}$-entropy, $\Sigma\mathbb{B}$-entropy, and $\Sigma\mathbf{T}$-entropy.

If one imposes the ordering on these four entropies, as in Sect.1.4, then one obtains the following picture; Fig.2.

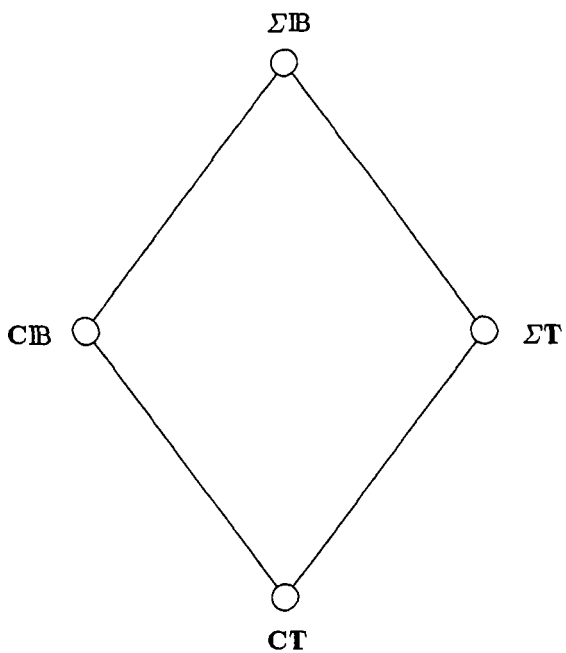


**Fig. 2.** The ordering of four entropies: $\Sigma\mathbb{B}$, $\mathbf{C}\mathbb{B}$, $\Sigma\mathbf{T}$, $\mathbf{CT}$

The four entropies of this section also admit definitions with slightly modified versions of conditions $\mathbf{C}$ and $\Sigma$.

**Condition C′.** There exists a constant $b$ such that the cardinality of $M$ is less

than or equal to $b \cdot 2^n$ for every $M \subseteq Y$ satisfying the requirements (i) and (ii) of Condition **C**.

**Condition $\Sigma'$.** There exists a constant $b$ such that

$$\sum_{y \in M} 2^{-f(y)} \leq b$$

for every $M \subseteq Y$ of mutually non-comparable elements.

Classes $\mathcal{Z}(\mathbf{C}', \mathbb{IB})$, $\mathcal{Z}(\mathbf{C}', \mathbf{T})$, $\mathcal{Z}(\Sigma', \mathbb{IB})$, and $\mathcal{Z}(\Sigma', \mathbf{T})$ differ from the corresponding classes $\mathcal{Z}(\mathbf{C}, \mathbb{IB})$, $\mathcal{Z}(\mathbf{C}, \mathbf{T})$, $\mathcal{Z}(\Sigma, \mathbb{IB})$, and $\mathcal{Z}(\Sigma, \mathbf{T})$; nevertheless, the related entropies coincide in the sense of Important Remark of Sect.1. That is,

$$\mathbf{C}'\mathbb{IB} = \mathbf{C}\mathbb{IB}, \ \mathbf{C}'\mathbf{T} = \mathbf{CT}, \ \Sigma'\mathbb{IB} = \Sigma\mathbb{IB}, \ \text{and} \ \Sigma'\mathbf{T} = \Sigma\mathbf{T}.$$

**Condition $\Sigma^\infty$.** For an arbitrary set $M \subseteq Y$ of mutually non-comparable elements,

$$\sum_{y \in M} 2^{-f(y)} < +\infty.$$

It is obvious that Condition $\Sigma^\infty$ is equivalent to Condition $\Sigma'$ for $Y = \mathbb{IB}$. Hence, the $\Sigma^\infty\mathbb{IB}$-entropy coincide with the $\Sigma'\mathbb{IB}$-entropy and consequently with the $\Sigma\mathbb{IB}$-entropy.

**Theorem (Andrei Muchnik).** *The conditions $\Sigma'$ and $\Sigma^\infty$ are equivalent in the case $Y = \mathbf{T}$.*

Hence the $\Sigma^\infty\mathbf{T}$-entropy coincides with the $\Sigma'\mathbf{T}$-entropy and with the $\Sigma\mathbf{T}$-entropy.

### 1.6 Relations between Two Quadruplets of Entropies

Now we have two quadruplets of entropies: the quadruplet $\mathbb{IBIB}$, $\mathbb{IBT}$, $\mathbf{T}\mathbb{IB}$, and $\mathbf{TT}$, which respectively generated by means of encoding, and the quadruplet $\mathbf{C}\mathbb{IB}$, $\mathbf{CT}$, $\Sigma\mathbb{IB}$, and $\Sigma\mathbf{T}$, which respectively generated by using some quantitative approach represented by conditions **C** and $\Sigma$.

It turns out that (in the sense of the equality explained in Sect.1, Important Remark) the following relations hold:

$$\mathbf{C}\mathbb{IB} = \mathbb{IBIB}, \ \mathbf{CT} = \mathbb{IBT}, \ \text{and} \ \Sigma\mathbb{IB} = \mathbf{T}\mathbb{IB}.$$

As to $\Sigma\mathbf{T}$, we have the following non-trivial fact, which will be discussed in Sect.2.2(5):

$$\Sigma\mathbf{T} < \mathbf{TT}.$$

Summarizing them, we obtain the following picture; Fig.3.

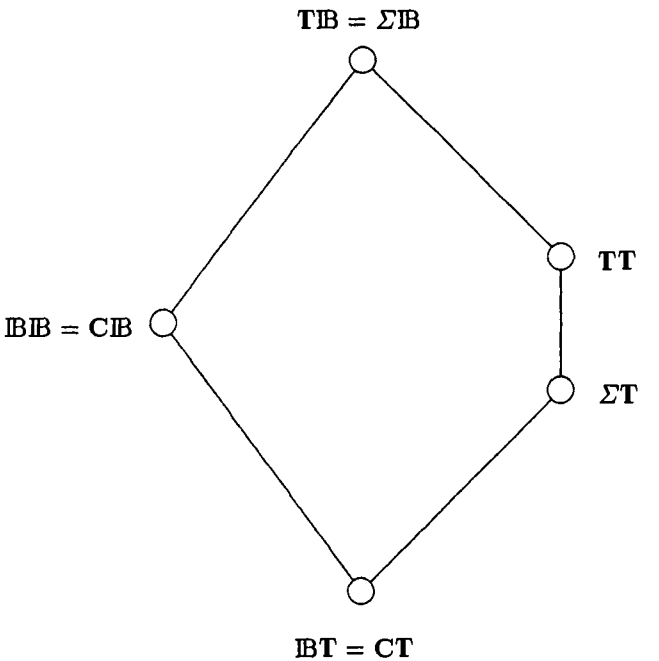


**Fig. 3.** The relation between entropies

### 1.7 A Semantic for ΣT-Entropy

Four entropies of Fig.3 have an encoding semantic, but the fifth entropy, $\Sigma\mathbf{T}$, has not yet obtained an appropriate semantic. Now a semantic for $\Sigma\mathbf{T}$ will be set forth. That semantic is based upon probabilistic machines.

To this end let us consider a probabilistic Turing machine with one-way infinite output tape whose head moves in only one direction. "Probabilistic" means that one must flip a symmetric coin before performing any command, and the result of flipping determines which command is to be performed. Another version: at the input tape, there comes a random infinite binary sequence with equal probabilities of digits. We suppose that our machine has binary output alphabet and never stops, so a finite or infinite binary sequence appears on the output tape.

Let us fix a machine $M$. For any $y \in \Xi$, let us denote by $\mathrm{P}^{\mathrm{n}}_M(y)$ the probability of the event '$y$ is the beginning of the output sequence'; in this notation, "n" stands for "non-stop". Consider a preorder relation $\leq$ on the set of machines: $M \leq N$ means that $\mathrm{P}^{\mathrm{n}}_M(y) \underset{\cap}{\leq} \mathrm{P}^{\mathrm{n}}_N(y)$.

**Explanation.** $A(y) \underset{\cap}{\leq} B(y)$ means that for some constant $c$ not depending on $y$, and for every $y$, $A(y) \leq c \cdot B(y)$.

It turns out that there exists a *maximal* machine $W$ such that $M \leq W$ for every machine $M$. In fact, there are several such machines, but any two of them, $U$ and $V$, satisfy the condition

$$\mathrm{P}^{\mathrm{n}}_U(y) \underset{\cap}{=} \mathrm{P}^{\mathrm{n}}_V(y).$$

**Explanation.** $A(y) \underset{\cap}{=} B(y)$ iff $A(y) \underset{\cap}{\leq} B(y)$ and $B(y) \underset{\cap}{\leq} A(y)$.

Hence for any two maximal machines $U$ and $V$, we have

$$|\log_2 \mathrm{P}^{\mathrm{n}}_U(y)| \underset{\cap}{=} |\log_2 \mathrm{P}^{\mathrm{n}}_V(y)|.$$

So we have moved from the probability $\mathrm{P}^{\mathrm{n}}_W$ to its logarithm. For any maximal machine $W$ one can verify that

$$|\log_2 \mathrm{P}^{\mathrm{n}}_W(y)| \underset{\cap}{=} \Sigma\mathbf{T}(y)$$

This fact enables us to identify $|\log_2 \mathrm{P}^{\mathrm{n}}_W|$ (or, if you prefer, the integer $\lfloor|\log_2 \mathrm{P}^{\mathrm{n}}_W|\rfloor$) with $\Sigma\mathbf{T}$. (Recall again Important Remark of Sect.1). Then the probabilistic definition of $\mathrm{P}^{\mathrm{n}}_W$ just given can be taken as a semantic for $\Sigma\mathbf{T}$.

The probability $\mathrm{P}^{\mathrm{n}}_W(y)$, related to an arbitrary maximal machine $W$, can be called *a priori probability* of $y$ as an element of the tree $\mathbf{T}$.

**Remark.** There exists also the a priori probability of $y$ as an element of the bunch $\mathbb{B}$. To obtain the a priori probability of that second sort, one should consider probabilistic machines of a slightly different type. The change is: instead of machines that never stop, one should take now probabilistic machines that can stop. Then $\mathrm{P}^{\mathrm{s}}_M(y)$ is, by definition, the probability of the event 'the word printed on output tape after machine $M$ stops coincides with $y$'; here "s" stands for "stop". A preorder on machines and the notion of a maximal machine are defined as above, and maximal machines do exist. Then $\mathrm{P}^{\mathrm{s}}_W(y)$ calculated for an arbitrary maximal machine $W$ is the a priori probability of $y$ as an element of $\mathbb{B}$. Here it occurs that

$$|\log_2 \mathrm{P}^{\mathrm{s}}_W(y)| \underset{\cap}{=} \Sigma\mathbb{B}(y).$$

Hence $\Sigma\mathbb{B}$ has a probabilistic semantic too. But, since $\Sigma\mathbb{B} = \mathbf{T}\mathbb{B}$, the entropy $\Sigma\mathbb{B}$ has also an encoding semantic.

## 1.8 Historical, Bibliographical, and Terminological Remarks; Acknowledgments

We begin the history of the theory of Kolmogorov complexity with Kolmogorov's paper [Kol65]. The purpose of that paper was to bring the notion of complexity (now we should say "of entropy") to the foundations of information theory. In his paper Kolmogorov expounded some results of his studies of 1963–1964. In those years he knew nothing about the paper [Sol64] in which Ray Solomonoff presented some similar ideas — but in vague and rather non-mathematical manner. We place the paper [Sol64] in the prehistory of the theory of Kolmogorov complexity. At the early stage of the theory's development, an important role belonged to the paper [ZL70].

In the papers of pioneers of the theory, there were introduced all five basic entropies of our Sect.1.6. The authors gave them various names and various notations. What was common in all those notations was the use of the letter "K" or the letter "k" as a part of the notation; one should believe the cause of this usage is a homage to Kolmogorov. Here we try to set some system of names and notations with the observance of the historical tradition. (In such a way the author makes his own contribution to the existing chaos of names and notations. This contribution is not too great because some names and notations are already in use. Simultaneously the author expresses the hope to introduce a standard system.)

We would like to fix the following names and notations for the five basic entropies.

1. For $\mathbb{IBIB}$-entropy. Name: **simple entropy**; notation: KS.
2. For $\mathbb{IBT}$-entropy. Name: **decision entropy**; notation: KD.
3. For $\mathbf{TIB}$-entropy. Name: **prefix entropy**; notation: KP.
4. For $\mathbf{TT}$-entropy. Name: **monotonic entropy**; notation: KM.
5. For $\Sigma\mathbf{T}$-entropy. Name: **a priori entropy**; notation: KA.

The entropies should be attributed to the following authors:

- simple entropy KS to Kolmogorov [Kol65](§3) and also (though in some nebulous form) to Solomonoff [Sol64],
- decision entropy KD to Loveland [Lov69],
- a priori entropy KA to Levin [ZL70](n° 3.3) and [Lev73],
- monotonic entropy KM to Levin [Lev73], and
- prefix entropy KP to Levin [Lev76].

**Remark.** Strictly speaking, we denote by the symbols KS, KD, KA, KM and KP exactly those versions of entropies as they were formulated by Kolmogorov, Loveland, and Levin. Let us recall the Important Remark of Sect.1. The coincidence KS with $\mathbb{IBIB}$ has the following meaning: for any particular entropy KS and for any particular entropy $\mathbb{IBIB}$, there holds $\mathrm{KS}(y) \underset{\Pi}{=} \mathbb{IBIB}(y)$. The other coincidences, KD with $\mathbb{IBT}$, etc., are to be understood in the same way.

Attributing the entropies to their inventors, we make no claim about the usage of these notations by the inventors. None of them made any essential use of the term "entropy"; usually the term "complexity" was used. Kolmogorov used simply the word "complexity" with no adjective. Loveland used the term "uniform complexity", and it was renamed as "decision complexity" by Zvonkin and Levin [ZL70](Definition 2.2). Levin used the words "monotonic complexity" and "complexity related to a prefix algorithm". He had not introduced any name for KA, but used terms "universal semicomputable measure" (in [ZL70](n° 3.3)) and "a priori probability" (in [Lev73]) for related quantities of which the logarithm is to be taken.

As to notations, Kolmogorov in [Kol65](§3) employed the notation $\mathrm{K}_A(y)$ for the simple entropy. Loveland in [Lov69](p.513) employed the notation $\mathrm{K}_A(x^n; n)$ for the decision entropy; and Zvonkin and Levin used for it the notation KR [ZL70](Definition 2.2). Zvonkin and Levin in [ZL70](n° 3.3) employed the notation $-\log_2 \mathrm{R}\{\Gamma_x\}$ for the a priori entropy; later, in [Lev73] that entropy was denoted by Levin as kM. In the same paper [Lev73] the notation km was used for the monotonic entropy. The notation KP (for the prefix entropy) appeared in [Lev76].

The general idea of an approximation space as a space of informations refining, or exactifying, one another is, without doubt, due to D. Scott. This idea was embodied into the notion of $f_0$-space in the sense of Yu. Ershov. A classification of entropies on the basis of that notion is given in [She84](Theorem 8); the classification of our Sect.1.3 is very close to that of [She84]. The general idea of the encoding-free approach to entropies (see Sect.1.5 above) was laid down in [Lev76].

A very useful exposition of various entropies and their interrelation is given in [Vyu81]. A survey of the use of entropies in a definition of randomness is presented in [KU87a] and [KU87b].

In the process of preparing this paper, the author had many discussions with Andrei Muchnik, Alexander Shen', and Nikolai Vereshchagin. The author enjoyed their advice and help. Many final formulations emerged from those thankworthy discussions. The bounds of Sect.2.1 and of Sect.2.2 probably belongs to what is called "mathematical folk-lore", but the final formulae are also due to discussions with Muchnik, Shen', and Vereshchagin.

To conclude this section let us redraw Fig.3 in terms and notations that we accept as standard; Fig.4.

The pentagon of Fig.4 shows, in particular, that neither KA $<$ KS nor KS $<$ KA. The exclamation note attached to an entropy means that the entropy can be used in the Kolmogorov definition of randomness.

## 2 Quantitative Analysis on Entropies

### 2.1 Bounds for Entropies

Some upper and lower bounds for entropies will be written down in this section. But first of all the reader must be warned that in this exposition, the sense of an

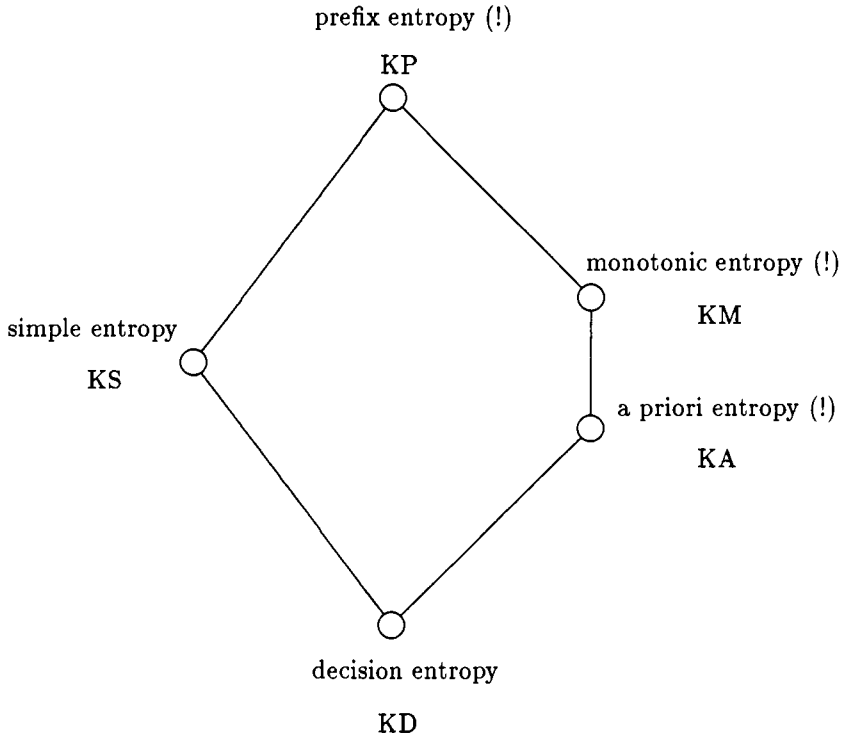


**Fig. 4.** Five basic entropies

upper bound and the sense of a lower bound are rather different. An upper bound for an entropy shows that the entropy cannot be too large. A lower bound for an entropy does not show that the entropy cannot be too small but does show that in infinitely many instances, the entropy can be large enough. So upper bounds are absolute, or strong, upper bounds. Lower bounds are not absolute; we shall call them *weak* lower bounds. A weak lower bound has the purpose of supporting the corresponding upper bound and demonstrating, in a favorable case, that the upper bound cannot be improved.

After this warning let us consider the five basic entropies of Sect.1.8, Fig.4.

(1) Entropies KS, KM, KA, and KD.

Let *Ent* denotes one of the entropies KS, KM, KA, KD. Then (an upper bound) for all $y$ (in $\Xi$),

$$Ent(y) \underset{+}{\leq} |y|,$$

and (a weak lower bound) for infinitely many $y$ (in $\Xi$),

$$Ent(y) \gtrsim |y|. \tag{1}$$

Let us formulate the lower bound of (1) more exactly. We have four cases: $Ent = \mathrm{KS}$, $Ent = \mathrm{KM}$, $Ent = \mathrm{KA}$, $Ent = \mathrm{KD}$. For each of these cases, the symbol $Ent$ (as well as KS, KM, KA or KD) denotes an arbitrary function belonging to some collection, i.e., the collection of $Ent$-entropies. In each case the meaning of (1) is as follows: for any particular function $Ent$ of that collection, there exist a constant $c$, perhaps negative, and an infinite set $M \subseteq \Xi$ such that

$$\forall y \in M \; [\; Ent(y) \geq |y| + c \;].$$

(2) Entropy KP.

It is helpful to introduce some notation. A function qlog, quasilogarithm, is introduced by the following definition:

$$\mathrm{qlog}z = \begin{cases} \log_2 z, & z \geq 1, \\ 0, & z \leq 1. \end{cases}$$

The iterations of that function are defined as follows:

$$\mathrm{qlog}^{(1)}z = \mathrm{qlog}z, \quad \text{and} \quad \mathrm{qlog}^{(k+1)}z = \mathrm{qlog}(\mathrm{qlog}^{(k)}z).$$

Then we have for any $k$, any $\epsilon > 0$, and all $y$,

$$\mathrm{KP}(y) \lesssim |y| + \mathrm{qlog}^{(1)}|y| + \mathrm{qlog}^{(2)}|y| + \cdots + \mathrm{qlog}^{(k-1)}|y| + (1+\epsilon)\mathrm{qlog}^{(k)}|y|. \tag{2}$$

It is an upper bound for KP.

For a weak lower bound, let $k$ be an arbitrary positive integer. Then, for infinitely many $y$,

$$[\; \mathrm{KP}(y) \geq |y| + \mathrm{qlog}^{(1)}|y| + \mathrm{qlog}^{(2)}|y| + \cdots + \mathrm{qlog}^{(k)}|y| \;]. \tag{3}$$

That means that inequality (3) holds for any function denoted by KP (i.e. for any prefix entropy) and for an appropriate infinite set of $y$'s, depending on the choice of that function.

Now it is reasonable to introduce an abbreviation for the sum $\mathrm{qlog}^{(1)}z + \cdots + (1+\epsilon)\mathrm{qlog}^{(k)}z$. Let us take, e.g., $Q_k(z,\epsilon)$ as such an abbreviation. That is,

$$Q_k(z,\epsilon) = \mathrm{qlog}^{(1)}z + \mathrm{qlog}^{(2)}z + \cdots + \mathrm{qlog}^{(k-1)}z + (1+\epsilon)\mathrm{qlog}^{(k)}z.$$

Then (2) and (3) can be rewritten as follows:

$$\forall y \; [\; \mathrm{KP}(y) \lesssim |y| + Q_k(|y|,\epsilon) \;],$$

and

$$\text{for infinitely many } y \; [\; \mathrm{KP}(y) \geq |y| + Q_k(|y|,0) \;].$$

## 2.2 Bounds for Differences of Entropies

By how much can two entropies of different sorts, e.g., KP and KD, can differ from one another? Perhaps it is better to ask, how much can one entropy exceed the other? Upper and lower bounds are to give the answer. The warning of the beginning of Sect.2.1 about the different meanings of upper and lower bounds is valid here and now too.

When $A(y) \underset{\oplus}{<} B(y)$, an upper bound for the difference $A(y) - B(y)$ is trivial; namely, a constant. So in this section, only differences $A(y) - B(y)$ for which the assertion $A(y) \underset{\oplus}{<} B(y)$ is false will be studied.

Now let us proceed to the differences.

(1) Difference KP − KD.
For any $k$, any $\epsilon > 0$, and all $y$,

$$\mathrm{KP}(y) - \mathrm{KD}(y) \underset{\oplus}{<} \mathrm{qlog}|y| + Q_k(|y|, \epsilon). \tag{4}$$

For any $k$ and infinitely many $y$,

$$\mathrm{KP}(y) - \mathrm{KD}(y) \geq \mathrm{qlog}|y| + Q_k(|y|, 0). \tag{5}$$

*Note.* Let us not forget that an additive constant implied in (4) depends not only on $k$ and $\epsilon$ but also on the particular versions of KP and KD. In (5) the set of y's depends not only on $k$ but also on the versions of KP and KD. This point is valid for further inequalities related to lower and upper bounds.

(2) Differences KS − KM and KS − KA.
Let $|y| \neq 0$. Then, for all $y$,

$$\mathrm{KS}(y) - \mathrm{KM}(y) \underset{\oplus}{<} \mathrm{KS}(y) - \mathrm{KA}(y) \underset{\oplus}{<} \log_2 |y|.$$

For some $c$ and for infinitely many $y$,

$$\mathrm{KS}(y) - \mathrm{KM}(y) \geq \log_2 |y| + c\,],$$

and for some $c$ and for infinitely many $y$,

$$\mathrm{KS}(y) - \mathrm{KA}(y) \geq \log_2 |y| + c.$$

(3) Differences KM − KS and KA − KS.
For any $k$, any $\epsilon > 0$, and all $y$,

$$\mathrm{KM}(y) - \mathrm{KS}(y) \underset{\oplus}{<} Q_k(|y|, \epsilon),$$
$$\mathrm{KA}(y) - \mathrm{KS}(y) \underset{\oplus}{<} Q_k(|y|, \epsilon).$$

For any $k$ and infinitely many $y$,

$$\mathrm{KM}(y) - \mathrm{KS}(y) \geq Q_k(|y|, 0),$$
$$\mathrm{KA}(y) - \mathrm{KS}(y) \geq Q_k(|y|, 0).$$

(4) Differences KP−KS, KS−KD, KP−KM, KP−KA, KM−KD, and KA−KD.

Let $B - A$ be any of the six entropy differences mentioned above. For any $k$, any $\epsilon > 0$, and all $y$,

$$B(y) - A(y) \underset{+}{\leq} Q_k(|y|, \epsilon).$$

And for any $k$ and infinitely many $y$,

$$B(y) - A(y) \geq Q_k(|y|, 0).$$

(5) The Difference KM − KA.

This difference is of special interest. The very fact that the conjecture $\mathrm{KM}(y) \underset{+}{=} \mathrm{KA}(y)$ is false is disappointing. The refutation of that conjecture is due to Petér Gács [Gac83]. (The Hungarian surname "Gács" is to be pronounced as English "garch".)

Both KM and KA are defined on the binary tree **T**. Gács studied two entropies K and H of similar sorts; but his K and H are defined not on **T** but on the tree consisting of all words in a countable alphabet (say, in $\mathbb{N}$ if one takes $\mathbb{N}$ as an alphabet). Some bound for the difference K − H is stated in Theorem 1.1 of [Gac83]; there the author writes: "Therefore for binary strings, the lower bound obtainable from the proof of Theorem 1.1 is only the inverse of some version of Ackermann's function" [Gac83](p.75). As it is known, Ackermann's function is a function from $\mathbb{N}$ to $\mathbb{N}$ which exceeds in its growth any primitive recursive function. The inverse $f^{-1}$ for a function $f$ is defined as follows:

$$f^{-1}(a) = \min\{z : f(z) \geq a\}.$$

Thus, for infinitely many $y$,

$$\mathrm{KM}(y) - \mathrm{KA}(y) \geq f^{-1}(|y|), \tag{6}$$

where $f$ is the version of Ackermann's function mentioned by Gács.

Let $\mathrm{Z}(y)$ denote the number of zeros in the word $y$. Then, as a corollary of Theorem 1.1 of [Gac83], we have the following: for any $k$ and for any $m$, there exists a $y \in \Xi$ such that $\mathrm{Z}(y) > m$ and

$$\mathrm{KM}(y) - \mathrm{KA}(y) \geq Q_k(Z(y), 0). \tag{7}$$

Therefore, we have two weak lower bounds. As to upper bounds, there is known no one except the following trivial one: for any $k$, any $\epsilon > 0$, and all $y$,

$$\mathrm{KM}(y) - \mathrm{KA}(y) \underset{+}{\leq} Q_k(|y|, \epsilon). \tag{8}$$

Since the weak lower bounds (6) and (7) do not support the upper bound (8), the task of improving all those bounds is open.

## References


[Gac83] P. Gács. On the relation between descriptional complexity and algorithmic probability. *Theoretical Computer Science* 22:71–93, 1983.

[Kol65] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems Inform. Transmission* 1:1–7, 1965. (Translated from the Russian version.)

[Kol68] A. N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. on Information Theory* IT-14.5:662–664, 1969. (The Russian version exists.)

[KU87a] A. N. Kolmogorov and V. A. Uspensky. Algorithms and randomness. In *Proc. 1st World Congress of the Bernoulli Society*, vol. 1. Yu. A. Prohorov and V. V. Sazonov, (eds.), VNU Science Press, Utrecht 3–53, 1987.

[KU87b] A. N. Kolmogorov and V. A. Uspensky. Algorithms and randomness. *Theory Probab. Appl.* 32:389–412, 1987. (Translated from the Russian version.) (*Comment:* There are two regrettable errors in the English version: p.394, line 2 from the bottom, and p.395, lines 1 and 3 from the top, the word "countable" must be replaced by "enumerable (i.e. recursively enumerable)"; p.395, line 1 from the top, the word "in" must be removed.)

[Lev73] L. A. Levin. On the notion of a random sequence. *Soviet Math. Dokl.* 14:1413–1416, 1973. (Translated from the Russian version.)

[Lev76] L. A. Levin. Various measures of complexity for finite objects (axiomatic description). *Soviet Math. Dokl.* 17:522–526, 1976. (Translated from the Russian version.)

[Lov69] D. W. Loveland. A variant of the Kolmogorov concept of complexity. *Information and Control* 15:510–526, 1969.

[Mar66] P. Martin-Lof. On the definition of random sequences. *Information and Control* 9:602–619, 1966.

[She84] A. Kh. Shen'. Algorithmic variants of the notion of entropy. *Soviet Math. Dokl.* 29:569–573, 1984. (Translated from the Russian version.) (*Comment:* There are many misprints in the English version.)

[Sol64] R. Solomonoff. A formal theory of inductive inference, Part I. *Information and Control* 7:1–22, 1964.

[US81] V. A. Uspensky and A. L. Semenov. What are the gains of the theory of algorithms: basic developments connected with the concept of algorithm and with its applications in mathematics (Part I,§17). In Springer-Verlag, Lecture Notes in Computer Science 122:100-234, 1981.

[US87] V. A. Uspensky and A. L. Semenov. *Teoria algoritmov: osnovnye otkrytiya i prilozheniya* (Theory of algorithms: main discoveries and applications) (§1.17). Nauka, Moscow, 1987; in Russian.

[Vyu81] V. V. V'yugin. Algorithmic entropy (complexity) of finite objects and its application to defining randomness and quantity of information. *Semiotika i Informatika* (Semiotics and Informatics) 16:14–43, 1981; in Russian.

[ZL70] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys* 25:83–124, 1970. (Translated from the Russian version.)

# Subject Index